# Proximal Convexification Procedures in Combinatorial Optimization

Aris Daniilidis, Claude Lemaréchal

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

# *Proximal Convexification Procedures in Combinatorial Optimization*

Aris Daniilidis, Claude Lemaréchal

**N° 4550**

Septembre 2002

THÈME 4

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4568--FR+ENG

# Proximal Convexification Procedures in Combinatorial Optimization

Aris Daniilidis, Claude Lemaréchal

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet Numopt

Rapport de recherche n° 4550 — Septembre 2002 — 18 pages

**Abstract:** Lagrangian relaxation is useful to bound the optimal value of a given optimization problem, and also to obtain relaxed solutions. To obtain primal solutions, it is conceivable to use a convexification procedure suggested by D.P. Bertsekas in 1979, based on the proximal algorithm in the primal space.

The present paper studies the theory assessing the approach in the framework of combinatorial optimization. Our results indicate that very little can be expected in theory, even though fairly good practical results have been obtained for the unit-commitment problem.

**Key-words:** Proximal algorithm, Lagrangian relaxation, duality gap, primal-dual heuristics

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 Montbonnot-St-Martin (France)
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

# Procédures Primales de Convexification en Optimisation Combinatoire

**Résumé :** La relaxation lagrangienne est utile pour borner la valeur optimale d'un problème d'optimisation, et aussi pour obtenir des solutions relaxées. Pour obtenir des solutions primales, on peut concevoir d'utiliser une procédure de convexification proposée par Bertsekas en 1979, fondée sur l'algorithme proximal dans l'espace primal.

Dans cet article, nous étudions la théorie sous-jacente à cette approche, dans le cadre de l'optimisation combinatoire. Nos résultats suggèrent que bien peu de justifications théoriques peuvent être attendues, même si de très bons résultats pratiques ont été obtenus, pour des problèmes d'optimisation de la production.

**Mots-clés :** Algorithme proximal, relaxation lagrangienne, saut dual, heuristiques primales-duales

# 1 Introduction, Motivation

This paper is motivated by a practical application: the unit-commitment problem, more precisely to optimize the generation schedules of the set of electrical power plants in France. Such a problem is usually solved through duality ([2, 11], see also [5, 12] for additional references). After solving the dual, comes the question of recovering a primal feasible solution, possibly suboptimal. An idea is to add in the production cost a quadratic term penalizing the deviation from the relaxed solution, obtained by dual means. This can give fairly good practical results, to be published elsewhere: [7].

Our aim here is to study theoretically this approach, with emphasis on combinatorial problems lending themselves to Lagrangian relaxation. To this aim, we consider first the general optimization problem

$$\inf f(x)\,, \quad x \in \mathbb{R}^n\,. \tag{1}$$

In this simplified notation, possible constraints are incorporated via the indicator function (0 on the feasible set, $+\infty$ outside). We have particularly in mind *0-1 linear programming problems*, say

$$\min b^\top x\,, \quad Ax \geqslant c \in \mathbb{R}^m\,, \quad x^i \in \{0,1\}\,,\; i = 1,\dots,n\,; \tag{2}$$

in this case the objective function is

$$f(x) := \begin{cases} b^\top x & \text{if } Ax \geqslant c \text{ and } x \in \{0,1\}^n\,, \\ +\infty & \text{otherwise}\,. \end{cases} \tag{3}$$

Our aim is to study for such problems a convexification procedure introduced by D.P. Bertsekas in [1], based on the (primal) *proximal algorithm*.

## 1.1 The General Idea: Moreau Envelope

The basic idea of this procedure is to add some convexity into $f$, which is replaced by

$$f_\rho(x,y) := f(x) + \rho\|x-y\|^2\,, \tag{4}$$

where $\rho > 0$ and $y \in \mathbb{R}^n$ is an additional variable. Then $f_\rho$ is minimized hierarchically: one sets

$$\varphi_\rho(y) := \inf\{f_\rho(x,y) : x \in \mathbb{R}^n\} \tag{5}$$

and one minimizes $\varphi_\rho$ *on the whole space* (observe that $\varphi_\rho$ is $+\infty$ nowhere – unless $f \equiv +\infty$). The function $\varphi_\rho$ thus defined is fairly known in nonlinear analysis. Introduced originally for a convex function $f$, it is usually called the Moreau-Yosida regularization; for more general situations see [13], where it is called the Moreau envelope.

**Lemma 1.1** *The following general properties hold.*

*(i) For all $y \in \mathbb{R}^n$, the function $\rho \mapsto \varphi_\rho(y)$ is nondecreasing.*

*(ii) $\varphi_\rho(y) \leqslant f(y)$ for all $y \in \mathbb{R}^n$.*

*Assume that $f$ is lower semicontinuous and bounded from below. Then:*

*(iii) if $y_0$ is a local minimum of $\varphi_\rho$ then $x = y_0$ is the unique minimum point of $f_\rho(\cdot, y_0)$ in (5). In particular $\varphi_\rho(y_0) = f(y_0)$;*

*(iv) local minima of $\varphi_\rho$ are also local minima of $\varphi_{\rho'}$ for $\rho' \geqslant \rho$.*

*Proof.* (i) Obviously, $f_\rho(x,y)$ in (4) is a nondecreasing function of $\rho$ and this property is transmitted to the infima.

(ii) Just observe that $\varphi_\rho(y) \leqslant f_\rho(y,y) = f(y)$.

(iii) Let $y = y_0$ in (5). Then our assumption implies that the infimum is attained at some $x_0$. Now for any $y$ close to $y_0$, we can write

$$f(x_0) + \rho\|y_0 - x_0\|^2 = \varphi_\rho(y_0) \leqslant \varphi_\rho(y) \leqslant f(x_0) + \rho\|y - x_0\|^2 ,$$

hence $\|y_0 - x_0\|^2 \leqslant \|y - x_0\|^2$. Take in particular $y = y_0 - t(y_0 - x_0)$, so that $y - x_0 = (1-t)(y_0 - x_0)$. Then $\|y_0 - x_0\|^2 \leqslant (1-t)^2\|y_0 - x_0\|^2$. Taking $t > 0$ small enough shows that $y_0 - x_0$ must be 0.

(iv) Using (i), we have for a local minimum $y_0$ of $\varphi_\rho$:

$$\varphi_{\rho'}(y) \geqslant \varphi_\rho(y) \geqslant \varphi_\rho(y_0) \quad \text{for } y \text{ close to } y_0 \,;$$

but from (iii) and (ii), $\varphi_\rho(y_0) = f(y_0) \geqslant \varphi_{\rho'}(y_0)$, which completes the proof. □

**Remark 1.2** *In these results, (iii) is the most important and its proof deserves comment. Being a min-function, $\varphi_\rho$ is usually not differentiable: the concept of derivative, or gradient, is then replaced by that of* directional derivatives*:*

$$\varphi'_\rho(y,d) := \lim_{t\downarrow 0} \frac{\varphi_\rho(y+td) - \varphi_\rho(y)}{t}, \qquad \textit{for given } d \in \mathbb{R}^n \;.$$

*Now a well-known formula (due to J.M. Danskin in [3]) says that, under appropriate assumptions on $f_\rho$, the directional derivative of functions given by (5) exists and has the expression*

$$\varphi'_\rho(y,d) = \min\left\{d^\top \nabla_y f_\rho(x,y) :\; x \textit{ minimizes } f_\rho(\cdot,y)\right\}. \tag{6}$$

*Here $\nabla_y f_\rho(x,y) = 2\rho(y-x)$ is the partial derivative of $f_\rho$ with respect to $y$. In plain words: when moving from $y$ to $y+td$ ($t > 0$ small), the marginal change of $\varphi_\rho$ is the* smallest *scalar product of $d$ with the partial derivatives of the minimand, computed at all the minimizing $x$'s. For a local minimum, this change must be nonnegative: $\varphi'_\rho(y,d) \geqslant 0$ for any $d \in \mathbb{R}^n$, i.e. $d^\top \nabla_y f_\rho(x,y) \geqslant 0$ for any minimizing $x$ and any $d \in \mathbb{R}^n$. This just means $\nabla_y f_\rho(x,y) = 2\rho(y-x) = 0$, i.e $x = y$ for any $x$ minimizing* (4)*.* □

## 1.2 Minimizing $f$ via $\varphi_\rho$

Intuitively, minimizing $\varphi_\rho$ in (5) is equivalent to minimizing $f$; this can be made precise:

**Theorem 1.3** *The minimization of $f$ and of $\varphi_\rho$ are related as follows:*

*(i)* $\inf\left\{f(x) : x \in \mathbb{R}^n\right\} = \inf\left\{\varphi_\rho(y) : y \in \mathbb{R}^n\right\}$.

*(ii) If $x^*$ minimizes $f$, then $x^*$ minimizes $\varphi_\rho$.*

*(iii) Assume $f$ is lower semicontinuous and bounded from below. If $y^*$ minimizes $\varphi_\rho$ then $y^*$ minimizes $f$.*

*Proof.* (i) For any $x$ and $y$ in $\mathbb{R}^n$, $f_\rho(x,y) \geq f(x)$; hence $\varphi_\rho(y) \geq \inf f$ and $\inf \varphi_\rho \geq \inf f$. On the other hand, Lemma 1.1 (ii) gives $\inf \varphi_\rho \leqslant \inf f$.

(ii) In view of (i) and of Lemma 1.1 (ii), $\inf \varphi_\rho = \inf f = f(x^*) \geqslant \varphi_\rho(x^*)$.

(iii) It follows from Lemma 1.1 (ii), (iii) that $f(y^*) = \varphi_\rho(y^*) \leqslant \varphi_\rho(y) \leqslant f(y)$ for all $y \in \mathbb{R}^n$. □

Thus, the approach replaces a single minimization (of $f$) by:
– the minimization of $f_\rho(\cdot, y)$, a function which is "more convex" than $f$,
– the minimization of $\varphi_\rho$.
Then a natural question is whether $\varphi_\rho$ has a chance of being convex; this essentially corresponds to $f$ being convex:

**Theorem 1.4** *The following holds, relating the convexity of $f$ with that of $\varphi_\rho$:*
*(i) Assume $f$ is bounded from below. Then $\varphi_\rho$ is convex whenever $f_\rho$ is convex (jointly with respect to $x$ and $y$);*
*(ii) $f_\rho$ is convex (jointly) if and only if $f$ is convex.*

*Proof.* (i) This is a classical result, see for example [9, Corollary B.2.4.5].

(ii) If $f$ is convex, then $f_\rho$ is obviously convex (jointly). If $f_\rho$ is convex, then in particular $x \mapsto f(x) = f_\rho(x,x)$ is convex. □

In view of Theorem 1.3, better convexification properties could hardly be expected from this procedure. Indeed (1) contains just about any optimization problem; in particular, there are instances of (1) which are difficult, but for which $f_\rho$ is convex in $x$ and computing $\varphi_\rho$ is "easy" – an example will be given in §3.4. In these cases, minimizing $f$ could not be equivalent to minimizing $\varphi_\rho$, if the latter were convex!

On the other hand, a classical convexification scheme is augmented Lagrangian. Applied to a problem such as (2), for example, it would add the term $\rho\|Ax - c\|^2$. This also corresponds to introducing a Moreau envelope, but in the *dual* space; it does result in a convex optimization problem for $\rho$ large enough but is hardly implementable; see [8, § XII.5.2] for example. As mentioned in [1], the present *primal* approach has the advantage of preserving separability of $f$, if any. The crucial point is that the quadratic term in (4) is a sum over the coordinates of the primal variable $x$; as such, it is not too complicating. In fact, we are interested in instances of (1) amenable to Lagrangian relaxation – such is the case of (2). Our aim will then be to reduce the duality gap, and/or to produce heuristic primal solutions, generated by the algorithm minimizing $\varphi_\rho$.

## 1.3 The Proximal Algorithm

The algorithm suggested in [1] to minimize $\varphi_\rho$ is essentially

$$y_{k+1} \in \operatorname{Argmin}\{f_\rho(x, y_k) : x \in \mathbb{R}^n\}, \tag{7}$$

where Argmin denotes the set of global minimizers, assumed nonempty (which is the case when $f$ is lower semicontinuous and bounded from below); naturally, the algorithm stops when $y_{k+1} = y_k$. This is called the *proximal algorithm*, whose convergence properties rely on the following result:

**Lemma 1.5** *Assume $y_{k+1}$ exists in* (7). *There holds at each iteration:*

$$f(y_{k+1}) \leqslant f(y_k) - \rho\|y_k - y_{k+1}\|^2 .$$

*As a result:*
– *either* $f(y_k) \to -\infty$,
– *or* $\sum_k \|y_{k+1} - y_k\|^2 < +\infty$.

*Proof.* By definition of $y_{k+1}$ and from Lemma 1.1 (ii),

$$f(y_{k+1}) + \rho\|y_{k+1} - y_k\|^2 = \varphi_\rho(y_k) \leqslant f(y_k) ,$$

which is the required inequality. In particular, the sequence $\{f(y_k)\}$ is decreasing:
– either $f(y_k) \to -\infty$,
– or $f(y_k)$ is bounded below; we obtain by summation

$$f(y_{K+1}) - f(y_1) \leqslant -\rho \sum_{k=1}^{K} \|y_{k+1} - y_k\|^2 ,$$

which shows that the series $\sum_k \|y_{k+1} - y_k\|^2$ converges. □

We explain the motivation of the proximal algorithm in the light of Remark 1.2. To avoid excessive generality, assume that $\varphi_\rho$ is a smooth function, namely that it has a gradient $\nabla\varphi_\rho(y)$ at every $y \in \mathbb{R}^n$. Then its directional derivatives are $\varphi'_\rho(y,d) = d^\top \nabla\varphi_\rho(y)$ for all $d \in \mathbb{R}^n$; with Danskin's formula (6), this clearly implies that $f_\rho(\cdot, y)$ must have a *unique* minimizer[1] $x(y)$, and that $\nabla\varphi_\rho(y) = 2\rho[y - x(y)]$. In particular, each next iterate $y_{k+1}$ in (7) exists and is defined without ambiguity. Besides, since

$$y_{k+1} = x(y_k) = y_k + \frac{1}{2\rho} 2\rho(x(y_k) - y_k) = y_k - \frac{1}{2\rho}\nabla\varphi_\rho(y_k) ,$$

we see that the proximal algorithm is just the minimization of $\varphi_\rho$ by a standard gradient method. Now, from Lemma 1.5:
– either $f(y_k) \to -\infty$ (then we are certainly minimizing $f$ successfully!)
– or $2\rho(y_k - y_{k+1}) = \nabla\varphi_\rho(y_k) \to 0$.
In the latter situation, assume that the sequence $\{y_k\}$ has some cluster point $y^*$. If $\varphi_\rho$ is actually continuously differentiable (this means that the mapping $y \mapsto x(y)$ is *continuous*), then we see that $\nabla\varphi_\rho(y^*) = 2\rho[y^* - x(y^*)] = 0$: the proximal algorithm can only produce stationary points of $\varphi_\rho$, which have a chance to be local minimizers.

# 2 Discrete Optimization Problems: Conceptual Forms

In this section we focus on the particular case where (1) is actually an optimization problem on a *finite* set: we consider

$$\min g(x) , \quad x \in F = \{x_1, \dots, x_K\} . \tag{8}$$

[1] To see this, observe that the directional derivative is linear in $d$, hence symmetric; with several minimizers, we would have $\varphi'_\rho(y,d) \neq -\varphi'_\rho(y,-d)$ for some direction $d$, a contradiction.

The function $g$ is left unspecified for the moment; it is completely characterized by its (finitely many) values $\{g(x_k)\}_{k=1}^K$. With respect to our previous notation,

$$f(x) = \begin{cases} g(x_k) & \text{if } x = x_k \text{ for some } k = 1, \ldots, K\,, \\ +\infty & \text{otherwise}\,. \end{cases} \tag{9}$$

Needless to say, this function is bounded and lower semi-continuous: the assumptions made in §1 are trivially satisfied. As for the "outer" objective function, it becomes

$$\varphi_\rho(y) = \min_{x \in F} \left\{ g(x) + \rho \|x - y\|^2 \right\}. \tag{10}$$

First of all, local minima of $\varphi_\rho$ can be characterized in this particular situation:

**Proposition 2.1** *A point $y_0$ is a local minimum of $\varphi_\rho$ if and only if $x = y_0$ is the unique optimal solution of* (10) *for $y = y_0$. In particular, every local minimum is a strict local minimum.*

*Proof.* In view of Lemma 1.1 (iii), we have only to prove that, if $x = y_0$ is the unique solution of (10) for $y = y_0$, then $y_0$ is a strict local minimum of $\varphi_\rho$. Note first that, since $F$ is a finite set, there is $\varepsilon > 0$ such that, for all $x \in F$ different from $y_0$,

$$g(x) + \rho \|x - y_0\|^2 \geqslant g(y_0) + 2\varepsilon = \varphi_\rho(y_0) + 2\varepsilon\,.$$

Now take $y$ close enough to $y_0$ so that, for all $x \in F$,

$$\rho \|x - y\|^2 \geqslant \rho \|x - y_0\|^2 - \varepsilon\,.$$

Summing these two inequalities, we obtain

$$g(x) + \rho \|x - y\|^2 \geqslant \varphi_\rho(y_0) + \varepsilon$$

and hence

$$\min \left\{ g(x) + \rho \|x - y\|^2 : x \in F \backslash \{y_0\} \right\} \geqslant \varphi_\rho(y_0) + \varepsilon\,.$$

As a result: for $y$ close enough to $y_0$,

$$\varphi_\rho(y) \geqslant \min \{ \varphi_\rho(y_0) + \varepsilon, \varphi_\rho(y_0) + \rho \|y_0 - y\|^2 \} \geqslant \varphi_\rho(y_0)\,,$$

and the second inequality is strict if $y \neq y_0$. □

Note that $\varphi_\rho$ is the minimum of finitely many quadratic functions. The $y$-space is divided into regions inside which the minimum in (10) is attained at some point $x \in F$, call it $x_k$ ($k$ depending on the region in question). The corresponding quadratic portion has the equation $g(x_k) + \rho \|y - x_k\|^2$, with gradient $2\rho(y - x_k)$. Altogether, $\varphi_\rho$ looks as indicated in Fig. 1. Local minima can be points such as $y_1$ or $y_2$; but at $y_3$, there are two minimum points in (10): $x = y_3$ and some other $x \in F$; in view of Lemma 1.1 (iii), $y_3$ cannot be a local minimum of $\varphi_\rho$. It is instructive to look at this picture with Remark 1.2 in mind.

We already know from Theorem 1.3 (iii) that local minima of $\varphi_\rho$ lie in $F$. Our next result specifies which feasible points can be thus obtained.

**Proposition 2.2** *For any $\rho > 0$, any local minimum of $\varphi_\rho$ lies in $F$. Moreover:*

*(i) For $\rho$ large enough, the local minima of $\varphi_\rho$ are exactly the points in $F$.*

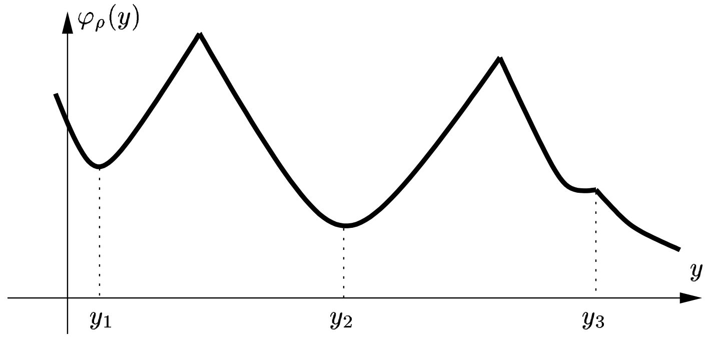


Figure 1: A piecewise quadratic function.

*(ii) For $\rho$ small enough, the local minima of $\varphi_\rho$ are exactly the optimal solutions of* (8).

*Proof.* If $y_0$ is a local minimum of $\varphi_\rho$, Proposition 2.1 shows that $y_0$ minimizes $f_\rho(x, y_0) = g(x) + \rho\|x - y_0\|^2$ over $x \in F$; then in particular $y_0 \in F$.
(i) We have to prove that an arbitrary $y_0 \in F$ is a local minimum of $\varphi_\rho$ if $\rho$ is large enough. Define the diameter of $g(F)$

$$\Gamma := \max\{g(x) - g(x') : x \in F, x' \in F\},$$

the "discreteness" of $F$

$$\varepsilon := \min\{\|x - x'\| : x \in F, x' \in F, x \neq x'\},$$

and take $\rho > \Gamma/\varepsilon^2$.

For $y$ close to $y_0$, namely for $\|y - y_0\| \leqslant \delta := \varepsilon - \sqrt{\Gamma/\rho} > 0$, there holds for all $x \in F$ different from $y_0$:

$$\|x - y\| \geqslant \|x - y_0\| - \|y_0 - y\| \geqslant \varepsilon - \delta\,.$$

Then we write

$$\begin{aligned} g(x) + \rho\|y - x\|^2 &\geqslant g(y_0) + g(x) - g(y_0) + \rho(\varepsilon - \delta)^2 \\ &\geqslant g(y_0) - \Gamma + \rho(\varepsilon - \delta)^2 \\ &= g(y_0) \geqslant \varphi_\rho(y_0)\,, \end{aligned}$$

where the last inequality is Lemma 1.1 (ii). The conclusion follows by taking the infimum over $x$ (knowing that the inequality holds trivially for $x = y_0$).
(ii) Call $v^*$ the optimal value in (8), $V^*$ the set of optimal solutions,

$$v^+ := \min\{g(x) : x \in F \backslash V^*\}$$

the "next to optimal" value of $g$ over $F$, and finally

$$D := \max\{\|x - x'\| : x, x' \in F\}$$

the diameter of $F$. Note that $v^+ > v^*$ and take $0 < \rho < (v^+ - v^*)/D^2$.

Take $x^* \in V^*$ and let $y$ be a local minimizer of $\varphi_\rho$. We already know that $y \in F$ and that $\varphi_\rho(y) = g(y)$ (Proposition 2.1); then write

$$g(y) = \varphi_\rho(y) \leqslant g(x^*) + \rho\|x^* - y\|^2 \leqslant g(x^*) + \rho D^2 < g(x^*) + v^+ - v^* = v^+\,.$$

By definition of $v^+$, $g(y)$ has to be equal to $v^*$; from Theorem 1.3 (i), this means $y \in V^*$. □

The proximal algorithm of §1.3 can then be described in the present context:

**Algorithm 2.3 (Conceptual proximal algorithm)** *Choose $\rho > 0$.*
STEP 0. *Take $y_1$ arbitrary in $\mathbb{R}^n$; set $k = 1$.*
STEP 1. *Let $x_k$ realize the smallest of the numbers $g(x_{k'}) + \rho\|y_k - x_{k'}\|^2$, for $k' = 1, \dots, K$.*
STEP 2. *If $x_k = y_k$ then stop.*
STEP 3. *Set $y_{k+1} = x_k$, replace $k$ by $k+1$ and go to Step 1.* □

Convergence is easy to establish:

**Theorem 2.4** *The above proximal algorithm terminates at some $k$ with a $y_k \in F$ which is a local minimum of $\varphi_{\rho'}$ for any $\rho' > \rho$.*

*Proof.* The $y_k$'s can take on finitely many values; in view of Lemma 1.5, $y_{k+1} - y_k$ tends to 0; hence $y_{k+1} - y_k = 0$ at some $k$. By construction, this means $x_k = y_k$ minimizes the function $x \mapsto g(x) + \rho\|x - y_k\|^2$:

$$\varphi_\rho(y_k) = g(y_k) \leqslant g(x) + \rho\|x - y_k\|^2 , \quad \text{for all } x \in F ;$$

the inequality becomes strict if $x \neq y_k$ and $\rho$ is replaced by $\rho' > \rho$. To finish, apply Proposition 2.1. □

Let us sum up this §2.

(i) First remember Lemma 1.1 (iv): when $\rho$ grows from 0 to $+\infty$, the local minima of $\varphi_\rho$ form nested sets, growing from the "ideal" set of optimal solutions of (8), to the "worst" whole feasible set. It is therefore advantageous to take a "small" $\rho$ (whatever this means).

(ii) The proximal algorithm produces such a local minimum – a feasible point for (8). In terms of the objective function $g$, the quality of this point depends on
- the initialization: in view of Lemma 1.5, the objective function is improved by at least $\rho\|y_{k+1} - y_k\|^2$ at each iteration; accordingly, a "good" point will be obtained if $y_1$ is itself "good";
- the value of $\rho$: in view of Proposition 2.2, only an optimal solution can be produced if $\rho$ is small enough.
- Note that, if we were able to guarantee a *global* minimum of $\varphi_\rho$, instead of local, then we would for sure have an optimal solution (Theorem 1.3).

Unfortunately, this algorithm is only conceptual anyway: $\varphi_\rho$ cannot be computed exactly – a fortiori minimized globally.

# 3 Discrete Optimization Problems: Implementable Forms

From now on, we assume the feasible set $F$ of (8) to be some structured set of 0-1 points; for convenience, we also assume $g$ to be linear. Thus we consider a frequent situation in combinatorial optimization:

$$\min b^\top x , \quad x \in F := \{x \in G : Ax \geqslant c \in \mathbb{R}^m\} , \quad G \subset \{0,1\}^n . \tag{11}$$

We may have $G = \{0,1\}^n$ but common instances have

$$G = \left\{x \in \{0,1\}^n : Bx \geqslant d \in \mathbb{R}^p\right\} . \tag{12}$$

The important feature is that $G$ is assumed to be an "easy polyhedron", in the sense that linear functions can "easily" be minimized over it; then $\varphi_\rho$ can be underestimated via Lagrangian relaxation.

## 3.1 Introducing Lagrangian Relaxation

Here, with the extra variable $\lambda \in \mathbb{R}^m_+$, we introduce the Lagrangian associated with (11) and the corresponding relaxation $\tilde\varphi_\rho$ of $\varphi_\rho$:

$$\begin{aligned} \tilde\varphi_\rho(y) &:= \sup_{\lambda \geqslant 0} \theta_\rho(\lambda, y)\,, \qquad \text{where} \\ \theta_\rho(\lambda, y) &:= \min_{x \in G} \left\{ b^\top x + \lambda^\top (c - Ax) + \rho \|x - y\|^2 \right\}. \end{aligned} \tag{13}$$

The function $\theta_\rho(\cdot, y)$ defined above is piecewise linear and its maximization is a standard convex optimization problem. On the other hand, computing $\theta_\rho(\lambda, y)$ (for given $\lambda$ and $y$) is "easy" because, using a standard trick in combinatorial optimization, the quadratic term $\|x-y\|^2$ can be "linearized"; see (18) below.

A crucial object in this approach is the bounded polyhedron obtained by convexifying $G$: we set

$$P := \{x \in \operatorname{co} G : Ax \geqslant c\}\,, \quad \tilde F := \operatorname{ext} P \tag{14}$$

(co denotes the convex hull, ext the set of extreme points). It is useful to understand that, being a set of 0-1 points, $F$ in (11) is also the set of extreme points in its own convex hull; $\tilde F$ is made up of $F$, plus some "parasitic" extreme points, which are all fractional. For more on Lagrangian relaxation, see for example [10] and the references therein.

We will need some more notation:

$$\ell_\rho(x, y) := b^\top x + \rho (e - 2y)^\top x\,, \quad g_\rho(x) := (b + \rho e)^\top x - \rho \|x\|^2\,, \tag{15}$$

where $e := (1, \ldots, 1)$ is the vector of all ones;

$$v^* := \min_{x \in F} b^\top x\,, \quad v_c := \sup_{\lambda \geqslant 0} \min_{x \in G} \left\{ b^\top x + \lambda^\top (c - Ax) \right\} \tag{16}$$

are respectively the optimal value of (11) and of its relaxation (naturally $v^* = \varphi_0(y)$ and $v_c = \tilde\varphi_0(y)$ for all $y$).

**Lemma 3.1** *The function $\rho \mapsto \tilde\varphi_\rho(y)$ is nondecreasing and $v_c \leqslant \tilde\varphi_\rho(y) \leqslant \varphi_\rho(y)$ for all $y$. It follows that*

$$v_c \leqslant \inf_{y \in \mathbb{R}^n} \tilde\varphi_\rho(y) \leqslant v^*.$$

*Proof.* Weak duality (see [10] for example) guarantees $\tilde\varphi_\rho(y) \leqslant \varphi_\rho(y)$. Furthermore, the Lagrangian $b^\top x + \lambda^\top (c - Ax) + \rho\|x - y\|^2$ in (13) is obviously a nondecreasing function of $\rho$; and this property is transmitted to the infima and suprema. The rest follows easily, use in particular (16) and Theorem 1.3 (i). □

The form (13) of $\tilde\varphi_\rho$, as well as the expression of $v_c$ in (16), are not easy to deal with, we express them via *minimization* problems:

**Proposition 3.2** *Use the notation* (15), (16). *There holds*

$$v_c = \min_{x \in P} b^\top x = \min_{x \in \tilde F} b^\top x \tag{17}$$

*and the function $\tilde{\varphi}_\rho$ has any of the following expressions:*

$$\begin{aligned} \tilde{\varphi}_\rho(y) &= \min_{x\in P} \ell_\rho(x,y) + \rho\|y\|^2 \\ &= \min_{x\in\tilde{F}} \ell_\rho(x,y) + \rho\|y\|^2 \,, \end{aligned} \tag{18}$$

$$\begin{aligned} \tilde{\varphi}_\rho(y) &= \min_{x\in P}[g_\rho(x) + \rho\|x-y\|^2] \\ &= \min_{x\in\tilde{F}}[g_\rho(x) + \rho\|x-y\|^2] \,. \end{aligned} \tag{19}$$

*Proof.* We prove (18) for $\rho \geqslant 0$; this will prove (19): in fact both problems have the same minimand. Then (17) will follow for $\rho = 0$.

Observe that $\|x\|^2 = e^\top x$ for $x \in G \subset \{0,1\}^n$. Then develop $\|y-x\|^2$ in (13) to realize that

$$\tilde{\varphi}_\rho(y) = \sup_{\lambda\geqslant 0} \min_{x\in G} \left\{\ell_\rho(x,y) + \lambda^\top(c - Ax)\right\} + \rho\|y\|^2 \,.$$

Then recognize duality applied to the problem

$$\min\{\ell_\rho(x,y) : x \in G, Ax \geqslant c\}\,.$$

This is classically equivalent to convexifying $G$ (see for example [4, 6, 12]). Finally, minimizing the linear function $\ell_\rho(\cdot,y)$ over the bounded polyhedron $P$ gives the same values if the minimization is restricted to the extreme points of $P$. □

**Remark 3.3** *Thanks to the extreme simplicity of the minimization problem in* (18) *(the set $\tilde{F}$ is finite), the function $\tilde{\varphi}_\rho$ is continuous; besides it increases at infinity (as fast as the squared norm). As a result, it does have some minimum point $y^*$.* □

To reproduce the results of Sections 1 and 2, we will use either the form (18) – which is simple enough –, or (19) – which has the general form (10), but with the substantial difference that the function $g$ depends on $\rho$.

**Proposition 3.4** *The point $y_0$ is a local minimum of $\tilde{\varphi}_\rho$ if and only if $x = y_0$ is the unique optimal solution in* (18), (19) *with $y = y_0$. This implies in particular that $y_0$ lies in $\tilde{F}$, is a strict local minimum of $\tilde{\varphi}_\rho$, and that*

$$\tilde{\varphi}_\rho(y_0) = \ell_\rho(y_0,y_0) + \rho\|y_0\|^2 = g_\rho(y_0)\,.$$

*Proof.* Apply Proposition 2.1 to the form (19) of $\tilde{\varphi}_\rho$. □

The lower bound obtained by minimizing $\tilde{\varphi}_\rho$ does improve the relaxed value $v_c$ of (16). We can even prove slightly more:

**Proposition 3.5** *For $\rho > 0$ small enough, any global minimizer $y_\rho$ of $\tilde{\varphi}_\rho$ also minimizes $x \mapsto b^\top x$ over $\tilde{F}$, and therefore satisfies $\tilde{\varphi}_\rho(y_\rho) = v_c + \rho\big(e^\top y_\rho - \|y_\rho\|^2\big)$.*

*As a result, if there is a duality gap, then*

$$v_c < \inf\tilde{\varphi}_\rho \leqslant v^*\,, \quad \textit{for all } \rho > 0\,.$$

*Proof.* For any $y \in \tilde F$, define the number

$$\varepsilon(y) := \inf\{\rho > 0 : y \text{ minimizes (globally) } \tilde\varphi_\rho\}\,.$$

This is a nonnegative number, possibly $+\infty$. Then define $\hat F := \{y \in \tilde F : \varepsilon(y) > 0\}$ and set

$$\varepsilon := \min\{\varepsilon(y) : y \in \hat F\}\,;$$

note that $\varepsilon > 0$ because $\hat F$ is a finite set.

Fix $\rho < \varepsilon$ and let $y_\rho$ be an arbitrary global minimizer of $\tilde\varphi_\rho$; clearly $y \in \tilde F \backslash \hat F$, so $\varepsilon(y_\rho) = 0$. This means that there exists a sequence $\rho_k \downarrow 0$ such that $y_\rho$ is a global minimizer of $\tilde\varphi_{\rho_k}$. In view of Proposition 3.4, $x = y_\rho$ minimizes $\ell_{\rho_k}(\cdot, y_\rho)$:

$$b^\top x + \rho_k(e - 2y_\rho)^\top x \geqslant b^\top y_\rho + \rho_k(e - 2y_\rho)^\top y_\rho \quad \text{for all } x \in \tilde F \text{ and } k = 1, 2, \ldots$$

Letting $\rho_k \downarrow 0$ shows that $y_\rho$ minimizes $x \mapsto b^\top x$ over $\tilde F$: $b^\top y_\rho = v_c$.

As a result, $\inf \tilde\varphi_\rho = \tilde\varphi_\rho(y_\rho) = v_c + \rho\big(e^\top y_\rho - \|y_\rho\|^2\big) \geqslant v_c$. If $\tilde\varphi_\rho(y_\rho)$ were equal to $v_c$, $y_\rho$ would be a 0-1 point, lying in $F$ and there would be no duality gap. The rest follows from monotonicity (Lemma 3.1). $\square$

**Proposition 3.6** *For $\rho$ large enough, the local minima of $\tilde\varphi_\rho$ contain the whole feasible set $F$ of* (11).

*Proof.* Take $\rho > |b|_\infty := \max_i |b_i|$ and let $y_0 \in F$; in particular, $y_0 \in \{0,1\}^n$. For any $x \in \tilde F \subset [0,1]^n$ and $i = 1, \ldots, n$, consider two cases:

- if $y_0^i = 0$, then $[b + \rho(e - 2y_0)]^i (x - y_0)^i = (b^i + \rho)(x^i) \geqslant 0$,
  the inequality being strict if $x^i > 0 = y_0^i$;
- if $y_0^i = 1$, then $[b + \rho(e - 2y_0)]^i (x - y_0)^i = (b^i - \rho)(x^i - 1) \geqslant 0$,
  the inequality being strict if $x^i < 1 = y_0^i$.

Thus, we see by summation that

$$\ell_\rho(x, y_0) - \ell_\rho(y_0, y_0) = [b + \rho(e - 2y_0)]^\top (x - y_0) \geqslant 0\,,$$

the inequality being strict if $x \neq y_0$. Because $y_0 \in F \subset \tilde F$, the only optimal solution in (18) is clearly $x = y_0$. The result follows from Proposition 3.4. $\square$

Finally we show that the upper bound $v^*$ is attained in Proposition 3.5.

**Proposition 3.7** *For $\rho$ large enough, the global minima of $\tilde\varphi_\rho$ are exactly the optimal solutions of* (11).

*Proof.* Let $\rho$ be so large that the local minima of $\tilde\varphi_\rho$ contain the whole of $F$ (Proposition 3.6). Consider first the local minima $y_0$ of $\tilde\varphi_\rho$ that lie in $F$. They satisfy (Proposition 3.4) $\tilde\varphi_\rho(y_0) = b^\top y_0 \geqslant v^*$, where equality holds exactly for those $y_0$'s solving (11).

Now set $\alpha := \min\big\{e^\top y - \|y\|^2 : y \in \tilde F \backslash F\big\}$; this is a positive number because $\tilde F \backslash F$ is made up of finitely many fractional points. Increase $\rho$ if necessary so that $\rho\alpha > v^* - v_c$ and let $y_0$ be a local minimum of $\tilde\varphi_\rho$ lying in $\tilde F \backslash F$. It satisfies

$$\tilde\varphi_\rho(y_0) = b^\top y_0 + \rho\big(e^\top y_0 - \|y_0\|^2\big) \geqslant b^\top y_0 + \rho\alpha > b^\top y_0 + v^* - v_c \geqslant v^*\,,$$

where the last inequality comes from (17). In view of Lemma 3.1, $y_0$ cannot be a global minimum of $\tilde\varphi_\rho$. $\square$

## 3.2 An Example

As an illustrative example, consider the problem

$$\min_{x\in F}\{x^1+4x^2\}\,,\quad \text{where } F:=\left\{x\in\{0,1\}^2 : x^1+2x^2\geqslant 2\right\}. \tag{20}$$

The feasible points are $x^*:=(0,1)$ (the optimal solution) and $e=(1,1)$, so that

$$\varphi_\rho(y)=\min\{4+\rho\|y-x^*\|^2, 5+\rho\|y-e\|^2\}\,.$$

Working out the calculations shows that

$$\varphi_\rho(y)=\begin{cases} 4+\rho\|y-x^*\|^2 & \text{if } (e-x^*)^\top y\leqslant \frac{1+\rho}{2\rho}\,,\\ 5+\rho\|y-e\|^2 & \text{otherwise}\,.\end{cases}$$

First observe that $(e-x^*)^\top x^*=0$ is always smaller than $(1+\rho)/2\rho$, hence $y=x^*$ is always a local minimum. Now consider three cases:

- If $(1+\rho)/2\rho<1$ (i.e. $\rho>1$) then $y=e$ is another local minimum – Proposition 2.2 (i).
- If $(1+\rho)/2\rho>1$ (i.e. $\rho<1$), this latter local minimum vanishes – Proposition 2.2 (ii).
- We leave it to the reader to check the case $(1+2\rho)/2\rho=1$, and in particular to see why $y=e$ is not a local minimum.

Now we study $\tilde\varphi_\rho$. The relaxed solution is $x_c=(1,1/2)$ (see (16)), with objective value $b^\top x_c=3$; $P$ is the triangle whose vertices make up $\tilde F=\{x^*,e,x_c\}$. Then

$$\tilde\varphi_\rho(y)=\min\{\ell_\rho(x^*,y),\ell_\rho(e,y),\ell_\rho(x_c,y)\}+\rho\|y\|^2\,,$$

where

$$\begin{aligned}\ell_\rho(x^*,y)&=4+\rho-2\rho y^2\,,\\ \ell_\rho(e,y)&=5+2\rho-2\rho(y^1+y^2)\,,\\ \ell_\rho(x_c,y)&=3+3\rho/2-\rho(2y^1+y^2)\,.\end{aligned}$$

Calculations are left to the reader. The final results are as indicated in Table 1. Note: $x^*$ becomes the global minimum when $3+\rho/4$ becomes larger than 4, i.e. when $\rho\geqslant 4$; compare with the theoretical results of Proposition 3.6, 3.7.

| $\rho$ | 0 | | 1/3 | | | 4 | | | +∞ |
|---|---|---|---|---|---|---|---|---|---|
| local minima | \| | $x_c$ | \| | $x_c$ $\quad$ $x^*$ | \| | $x_c$ | $x^*$ | $e$ | |
| $\tilde\varphi_\rho$ values | \| | $3+\rho/4$ | \| | $3+\rho/4$ $\quad$ 4 | \| | $3+\rho/4$ | 4 | 5 | |

Table 1: Local minima of $\tilde\varphi_\rho$.

## 3.3 The Relaxed Proximal Algorithm

Consider now the proximal algorithm (7) to minimize $\tilde\varphi_\rho$. It needs a black box to compute $\tilde\varphi_\rho(y)$ for a given $y\in\mathbb{R}^n$. Of course this is done by some optimization process, which produces an $x(y)$ solving one of the "equivalent" problems (13), (18) or (19).

Then we do the following.

**Algorithm 3.8 (Implementable proximal algorithm I)** *A black box is assumed available to compute $x(y)$ for given $y$. Choose $\rho>0$.*

STEP 0. *Take $y_1$ arbitrary in $\mathbb{R}^n$; set $k = 1$.*
STEP 1. *Call the black box to obtain $x(y_k)$.*
STEP 2. *If $x(y_k) = y_k$ then stop.*
STEP 3. *Set $y_{k+1} = x(y_k)$, replace $k$ by $k+1$ and go to Step 1.* □

**Theorem 3.9** *Assume that $x(y_k) \in \tilde{F}$ at each iteration $k$. Then the stop in Step 2 occurs at some finite iteration $K$.*

*Proof.* By construction, the whole sequence $\{y_k\}$ lies in $\tilde{F}$, a finite set. Applying Lemma 1.5 to the form (19), we see that the sequence has to be finite. □

**Remark 3.10** *This result outlines some weaknesses of Algorithm 3.8.*

*First, little can be said about the output $y_K$: all we know is that $y_K \in \operatorname{Argmin} \ell_\rho(\cdot, y_K)$ – which does not imply that $y_K$ is a local minimum of $\tilde{\varphi}_\rho$.*

*Besides, the assumption $x(y_k) \in \tilde{F}$ is rather dary:*

- *either we solve* (13) *by some dual algorithm; this produces a point in $P$, which has no reason to lie in $\tilde{F}$;*
- *or we solve* (18) *by the simplex algorithm; but this supposes a close description of the polyhedron $P$ – which, in practice, amounts to assuming $G = \{0,1\}^n$.*

□

As a conclusion, let us compare with the situation in §2.

(i) Replacing $\varphi_\rho$ by $\tilde{\varphi}_\rho$ has a substantial price. Instead of producing a feasible point for sure (Theorem 2.4), we may land at some parasitic point in $\tilde{F} \backslash F$, including the relaxed solution $x_c$ from (16).

(ii) Taking a small $\rho$, which was recommended, is now unwise, since this will probably produce the relaxed solution $x_c$ (Proposition 3.5). Note at this point that $y_1 = x_c$ is probably the most natural initialization in Algorithm 3.8.

(iii) The only way to escape from $x_c$ is to increase $\rho$. However this is dangerous, since the algorithm might produce a bad feasible point of (11) (Proposition 3.6); besides, it is not sure that larger values of $\rho$ will eventually eliminate $x_c$ from the set of local minima.

Note of course that rounding $x_c$ may not produce any optimal point; it may not even produce any feasible point (replace in §3.2 the inequality constraint by the equality $x^1 + 2x^2 = 2$).

(iv) Once again, global minimization of $\tilde{\varphi}_\rho$ would solve the problem; yet we should also take $\rho$ large enough (Proposition 3.7).

## 3.4 Penalizing the 0-1 Constraints

The trouble with the previous convexification procedure is that the quadratic term can by no means convexify the objective function $f$ of (9). In fact, the original motivation for [1] was to treat "ordinary" (continuous) nonlinear programming problems, whose Lagrangian had a Hessian which could be made positive definite. To put (11) into this mould, we introduce the penalty factor

$$p(x) := e^\top x - \|x\|^2 , \tag{21}$$

we take a large parameter $\kappa > 0$, and we consider problems of the type

$$\min \left[b^\top x + \kappa p(x)\right], \quad x \in Q \subset [0,1]^n, \tag{22}$$

$Q$ being a convex polyhedron. To define $Q$, we may for example just change $\{0,1\}$ to $[0,1]$ in the definition of $F$; it is generally accepted that this does not change (11). We give a proof of a slightly more general result:

**Theorem 3.11** *Assume that the convex polyhedron $Q$ in* (22) *contains the feasible set $F$ of* (11)*, but does not contain any other 0-1 point than those in $F$.*

*(i) Any optimal solution of* (22) *lying in $F$ is also optimal for* (11).

*(ii) For $\kappa$ large enough, the sets of optimal solutions in* (11) *and* (22) *coincide.*

*Proof.* (i) Just observe that (22) is a *relaxation* of (11): their objective functions coincide on the feasible set $F \subset \{0,1\}^n$ of the latter.

(ii) Use the notation (16). Because $p \geqslant 0$ on $Q \subset [0,1]^n$, $v^*$ is an upper bound for the optimal value of (22). Because $p$ is strictly concave on $Q \subset [0,1]^n$, the feasible set in (22) can be restricted to the set $\operatorname{ext} Q$ of its extreme points. Denote by $\bar{F} := \operatorname{ext} Q \backslash F$ the set of parasitic points. Set $\delta := \min_{x \in \bar{F}} p(x)$ and note that $\delta > 0$ because $p > 0$ on the finite set $\bar{F}$.

Take $\kappa > \dfrac{(v^* - v_c)}{\delta}$, let $x^* \in \operatorname{ext} Q$ solve (22) and assume $x^* \in \bar{F}$:

$$b^\top x^* + \kappa p(x^*) \geqslant v_c + \kappa\delta > v^*,$$

a contradiction. Thus, any $x^*$ optimal in (22) lies in $F$; in view of (i), $x^*$ solves (11). Besides, the optimal value of (22) is $v^*$. Conversely, any $\hat{x}$ optimal in (11) is feasible in (22) and has $b^\top \hat{x} + \kappa p(\hat{x}) = b^\top \hat{x} = v^*$: $\hat{x}$ is optimal in (22). □

Thus, the proximal convexification procedure can be applied to (22) as well: we define

$$\psi_{\rho k}(y) := \min_{x \in Q} \left[b^\top x + \kappa p(x) + \rho \|y - x\|^2\right] \tag{23}$$

and all the results of §1 apply, $\psi_{\rho k}$ playing the role of $\varphi_\rho$. Observe that the minimand $f_\rho(\cdot, y)$ in (23) is convex for $\rho \geqslant \kappa$: in contrast with $\varphi_\rho$, the computation of $\psi_{\rho k}$ is straightforward. Its global minimization is not straightforward, though: with relation to Theorem 1.4, the Hessian of $f_\rho(\cdot,\cdot)$ is $2\begin{pmatrix} (\rho-\kappa)I & -\rho I \\ -\rho I & \rho I \end{pmatrix}$ ($I$ is the identity in $\mathbb{R}^n$), which is never positive semidefinite.

**Algorithm 3.12 (Implementable proximal algorithm II)** *A black box is assumed available to solve* (23) *for a given $y$. Take $\rho > \kappa$.*

STEP 0. *Take $y_1$ arbitrary in $\mathbb{R}^n$; set $k = 1$.*

STEP 1. *Call the black box to obtain $x(y_k) \in Q$.*

STEP 2. *If $x(y_k) = y_k$ then stop.*

STEP 3. *Set $y_{k+1} = x(y_k)$, replace $k$ by $k+1$ and go to Step 1.* □

Of course, the choice $\rho > \kappa$ implies that the minimum in (23) is unique: $x(y_k)$ is well-defined in Step 2. Another reason for this choice will become apparent in Proposition 3.15 below.

**Theorem 3.13** *The sequence $y_k$ generated by the above algorithm has some cluster point, and any such cluster point is a local minimum $y^*$ of $\psi_{\rho k}$ on $Q$.*

*Proof.* The first statement holds because $y_k$ varies in $Q$, a compact set. Take a subsequence – also denoted by $y_k$ – such that $y_k \to y^*$. Because of Lemma 1.5, $x(y_k) - y_k \to 0$, hence $x(y_k) \to y^*$. Then pass to the limit in the inequality

$$b^\top x + \kappa p(x) + \rho\|x - y_k\|^2 \geqslant b^\top x(y_k) + \kappa p(x(y_k)) + \rho\|x(y_k) - y_k\|^2 \quad \text{for all } x \in Q$$

to see that $x = y^*$ is the unique minimum point in (23) for $y = y^*$. The result follows from Proposition 2.1. □

**Remark 3.14** *Theorem 3.11 allows a large choice for the feasible set $Q$ in* (22).
- *The most natural is $Q = P$ of* (14)*; however, this requires a description of* $\operatorname{co} G$*, which may not be easy numerically.*
- *When $G$ has the form* (12)*, one may also take the ordinary LP relaxation:*

$$Q = \left\{x \in [0,1]^n : Ax \geqslant c, Bx \geqslant d\right\};$$

*then* (23) *is an ordinary linear-quadratic program. However, this $Q$ is then larger than $P$ and may introduce more parasitic local minima.* □

The minimand of (23) is linear in $x$ when $\rho = \kappa$, and this turns out to reproduce §3.1:

**Proposition 3.15** *Suppose $Q$ in* (22) *is $P$ of* (14).
*Then $\psi_{\rho\rho}(y) = \tilde\varphi_\rho(y)$ for all $y \in \mathbb{R}^n$ and $\rho > 0$. It follows that the local minima of $\tilde\varphi_\rho$ are also local minima of $\psi_{\rho'\rho}$, for any $\rho' \geqslant \rho$.*

*Proof.* For all $x, y$,

$$b^\top x + \rho p(x) + \rho\|x - y\|^2 = b^\top x + \rho(e - 2y)^\top x + \rho\|y\|^2,$$

in which we recognize the respective minimands in (23) (with $\kappa = \rho$) and in (18). Minimizing them gives the same value.

Now invoke Lemma 1.1 (iv): the set of local minima of $\psi_{\rho k}$ can only increase when $\rho$ increases. □

Thus, the additional flexibility yielded by $\rho > \kappa$ improves nothing: this can only enlarge the set of local minima. The proximal algorithm has more chances to produce a parasitic point. A small bonus is obtained with respect to Algorithm 3.8, though: if the algorithm stops at Step 2, the corresponding $y_k$ is *for sure* a local minimum of $\psi_{\rho k}$ (Theorem 3.13).

**General Conclusion** We have studied in the framework of combinatorial optimization a general convexification procedure (the primal proximal algorithm), assessed by a useful heuristic for some special problem (unit-commitment). This procedure gives birth to various conceptual and implementable heuristic algorithms to generate primal solutions, and our results suggest that little is to be expected in theory from the approach.

First it should be mentioned that the assumptions allowing our study of implementable algorithms (§3) do not fit easily with the actual form of the unit-commitment problem: thermal plants and hydro valleys involve fairly more sophisticated models than (11).

More importantly, the ability itself of the approach in producing feasible suboptimal solutions is controversial. From our results in §3.1, such an ability implies a delicate tuning of the proximity parameter, to avoid

*Charybdis:* staying with a small $\rho$ at the relaxed solution $x_c$,

as well as

*Scylla:* jumping with a large $\rho$ to an uncontrollable feasible point in (8).

Yet, even if these two dangers are avoided, the mere production of a feasible point is never guaranteed; see again our discussion at the end of §3.1.

# References

[1] D.P. Bertsekas. Convexification procedures and decomposition methods for nonconvex optimization problems. *Journal of Optimization Theory and Applications*, 29:169–197, 1979.

[2] D.P. Bertsekas, G.S. Lauer, N.R. Sandell, and T.A. Posberg. Optimal short-term scheduling of large-scale power systems. *IEEE Transactions on Automatic Control*, AC-28:1–11, 1983.

[3] J.M. Danskin. The theory of max-min with applications. *SIAM Journal on Applied Mathematics*, 14(4):641–655, 1966.

[4] J.E. Falk. Lagrange multipliers and nonconvex programs. *SIAM Journal on Control*, 7(4):534–545, 1969.

[5] S. Feltenmark and K. C. Kiwiel. Dual applications of proximal bundle methods, including lagrangian relaxation of nonconvex problems. *SIAM Journal on Optimization*, 10(3):697–721, 2000.

[6] A.M. Geoffrion. Lagrangean relaxation for integer programming. *Mathematical Programming Study*, 2:82–114, 1974.

[7] R. Gonzalez and C. Lemaréchal. Unit-commitment problem: from the dual to the primal. In preparation.

[8] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms.* Springer Verlag, Heidelberg, 1993. Two volumes.

[9] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis.* Springer Verlag, Heidelberg, 2001.

[10] C. Lemaréchal. Lagrangian relaxation. In M. Jünger and D. Naddef, editors, *Computational Combinatorial Optimization*, pages 115–160. Springer Verlag, Heidelberg, 2001.

[11] C. Lemaréchal, F. Pellegrino, A. Renaud, and C. Sagastizábal. Bundle methods applied to the unit-commitment problem. In J. Doležal and J. Fidler, editors, *System Modelling and Optimization*, pages 395–402, 1996.

[12] C. Lemaréchal and A. Renaud. A geometric study of duality gaps, with applications. *Mathematical Programming*, 90(3):399–427, 2001.

[13] R.T. Rockafellar and R.J.-B. Wets. *Variational Analysis.* Springer Verlag, Heidelberg, 1998.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399